



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/703,174	10/31/2000	Charu C. Aggarwal	YOR920000430US1	7445
7590	06/28/2006		EXAMINER HILLERY, NATHAN	
William E Lewis Ryan Mason & Lewis LLP 90 Forest Avenue Locust Valley, NY 11560			ART UNIT 2176	PAPER NUMBER

DATE MAILED: 06/28/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

<b>Office Action Summary</b>	<b>Application No.</b> 09/703,174	<b>Applicant(s)</b> AGGARWAL ET AL.	
	<b>Examiner</b> Nathan Hillery	<b>Art Unit</b> 2176	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

### Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

### Status

- 1) ☒ Responsive to communication(s) filed on 27 April 2006.
- 2a) ☐ This action is **FINAL**.                      2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

### Disposition of Claims

- 4) ☒ Claim(s) 1-27 is/are pending in the application.
- 4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.
- 5) ☐ Claim(s) \_\_\_\_\_ is/are allowed.
- 6) ☒ Claim(s) 1-27 is/are rejected.
- 7) ☐ Claim(s) \_\_\_\_\_ is/are objected to.
- 8) ☐ Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

### Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on \_\_\_\_\_ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.  
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

### Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All    b) ☐ Some \*    c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
2. ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.
3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

\* See the attached detailed Office action for a list of the certified copies not received.

### Attachment(s)

- |  |   |
|--|---|
| 1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892)  | 4) <input type="checkbox"/> Interview Summary (PTO-413)<br>Paper No(s)/Mail Date. _____ |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948)                                   | 5) <input type="checkbox"/> Notice of Informal Patent Application (PTO-152)             |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)<br>Paper No(s)/Mail Date _____ | 6) <input type="checkbox"/> Other: _____  |

### DETAILED ACTION

1. This action is responsive to communications: RCE filed on 4/27/06.
2. Claims 1 – 27 are pending in the case. Claims 1, 10, and 19 are independent.

### ***Continued Examination Under 37 CFR 1.114***

3. A request for continued examination under 37 CFR 1.114, including the fee set forth in 37 CFR 1.17(e), was filed in this application after final rejection. Since this application is eligible for continued examination under 37 CFR 1.114, and the fee set forth in 37 CFR 1.17(e) has been timely paid, the finality of the previous Office action has been withdrawn pursuant to 37 CFR 1.114. Applicant's submission filed on 3/27/06 has been entered.

### ***Claim Rejections - 35 USC § 112***

4. The following is a quotation of the first paragraph of 35 U.S.C. 112:

The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same and shall set forth the best mode contemplated by the inventor of carrying out his invention.

5. Claims 1 – 27 are rejected under 35 U.S.C. 112, first paragraph, because the specification, while being enabling for *not assuming a specific model for web linkage structure* (Specification, p 4, line 24), does not reasonably provide enablement for *without using an initial link structure* (representatively - claim 1, lines 4 & 5). The specification does not enable any person skilled in the art to which it pertains, or with

which it is most nearly connected, to make and use the invention commensurate in scope with these claims.

6. It should be noted that although Applicant argues on p 7, last paragraph that support for the amendment, wherein **the initial document retrieval operation is performed without use of an initial link structure** in lines 3 – 5 of claim 1, is provided by not assuming a model of link structure (Specification, p 4, line 24), the Office does not interpret not assuming a model of link structure to be equivalent to not using a link structure at all. The World Wide Web or information network is a link structure.

However, the particular model of that link structure is something totally different.

7. Similarly, claims 2 – 27 are deemed to not be enabled for similar reasons.

8. Representative claim 1 recites “initially retrieving one or more documents from the information network that satisfy a user-defined predicate, wherein the initial document retrieval operation is performed without use of an initial link structure” in lines 3 – 5. Clearly, based on the above claim language, an information network is an initial link structure that is used to initially retrieve the documents. Consequently, the claim appears to be inoperative because the “initially retrieving one or more documents” has to be completed by using “an initial link structure” such as the claimed information network; therefore, it is unclear how “the initial document retrieval operation is performed without the use of an initial link structure”.

9. As is further evidenced by the cited NPL reference of Hofmann, by definition an information network is a link structure, since Hoffman teaches that *In the World Wide Web, myriads of hyperlinks connect documents and pages to create an unprecedented,*

*highly complex graph structure - the Web graph* (p 369, left column, lines 1 – 3).

Consequently, the claim limitation, “the initial document retrieval operation is performed without use of an initial link structure”, is interpreted as **the initial document retrieval operation is performed with use of an initial link structure**, since an information network is being interpreted as the “initial link structure” that is used to perform the “initial document retrieval operation” in contradistinction to the claim.

10. Similarly, claims 2 – 27 are also rejected for similar reasons.

11. The following is a quotation of the second paragraph of 35 U.S.C. 112:

The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

12. Claims 1 – 27 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

13. **Regarding independent claims 1, 10, and 19**, it is unclear what Applicant means by “initially retrieving one or more documents from the information network that satisfy a user-defined predicate, wherein the initial document retrieval operation is performed without use of an initial link structure”, since it appears to be inoperative and impossible. Consequently, the metes and bounds of the claim are unclear.

14. **Regarding dependent claims 2 – 9, 11 – 18 and 20 – 27**, the claims are rejected for fully incorporate the deficiencies of the base claim(s) from which they depend.

***Claim Rejections - 35 USC § 101***

15. 35 U.S.C. 101 reads as follows:

Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

16. Claims 1 – 27 are rejected under 35 U.S.C. 101 because the claimed invention is directed to non-statutory subject matter. Claims 1 – 27 have no practical application as claimed because there is no physical transformation and no production of a concrete, useful and tangible result.

- a. The result of the claimed invention remains in the abstract and is not made available to the user; thus it is not tangible.
- b. The claims appear to be in the preliminary stages and fall short of the disclosed practical utility. In other words, the claims fail to fulfill and/or reflect the specific, substantial, and credible utility sought by the disclosed invention, and thus do not produce a useful result.
- c. The input retrieved and collected by the invention appear to be subjectively analyzed with no reliable, assured result being output, and thus does not produce a concrete result.

17. Consequently, the claims are nonstatutory. The claims simply recite retrieving and collecting data and/or information with no concrete, useful, tangible result.

18. Further, to expedite a complete examination of the instant application the claims rejected under 35 U.S.C. 101 (nonstatutory) above are further rejected as set forth below in anticipation of applicant amending these claims to place them within the four statutory categories of invention.

***Claim Rejections - 35 USC § 102***

19. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

20. Claims 1 – 8, 10 – 17 and 19 – 26 are rejected under 35 U.S.C. 102(b) as being anticipated by Chakrabarti et al. (Focused Crawling: A New Approach to Topic-specific Web Resource Discovery) [as cited by applicant].

21. **Regarding independent claim 1**, Chakrabarti et al. teach that *keyword search is used to locate an initial set of pages (using a giant crawl and index)* (p 6, section 2.2, last paragraph), which is equivalent to the claimed **initially retrieving one or more documents from the information network that satisfy a user-defined predicate**. It should be noted that the *keyword search* of Chakrabati et al. is equivalent to the claimed **user-defined predicate**.

22. Chakrabarti et al. teach that while fetching a document, the above formulation is used to find the leaf node with the highest probability. If some ancestor has been marked good we allow future visitation of URLs found on the document, otherwise the crawl is pruned there (p 9, section Hard focus rule), which is equivalent to the claimed **collecting statistical information about the one or more retrieved documents as the one or more retrieved documents are analyzed and using the collected statistical information to automatically determine further document retrieval operations**, since the probabilities are calculated to find the “best” leaf node, the

ancestors are analyzed to determine if they are good, and then based on that finding future visitations are allowed (p 9, section Hard focus rule). It should be noted that the *probabilities* of Chakrabarti et al. are equivalent to the claimed **statistical information**.

23. Chakrabarti et al. teach that a focused crawler is an example-driven automatic porthole-generator. We feel that the ability to focus on a topical subgraph of the Web, as in this paper, together with the ability to browse communities within that subgraph, will lead to significantly improved Web resource discovery (p 3, last paragraph before Section 2), which is equivalent to the claimed **wherein the statistical information-using step further comprises learning a link structure from at least a portion of the collected statistical information with each successive document retrieval operation**. It should be noted that the *porthole*, which is a *subgraph of the Web*, generated by the *focused crawler* of Chakrabarti et al. is equivalent to the claimed **link structure** that is learned. It should further be noted that the generation of a porthole or specialized link structure (p 20, last paragraph) is equivalent to the claimed **learning a link structure**.

24. **Regarding dependent claim 2**, Chakrabarti et al. teach that Query construction is not a one-time investment, because as pages on the topic are discovered, their additional vocabulary must be folded in manually into the query for continued discovery (p 7, lines 4 – 6), which is equivalent to the claimed **the user-defined predicate specifies content associated with a document**. It should be noted that the *additional*



*vocabulary of pages on the topic* of Chakrabati et al. is equivalent to the claimed **content associated with a document**.

25. **Regarding dependent claims 3 and 4**, Chakrabarti et al. teach that pages that are examples associated with a topic can be preprocessed as desired by the system. The user's interest is characterized by a subset of topics that is marked good. No good topic is an ancestor of another good topic. Ancestors of good topics are called path topics. Given a Web page, a measure of its relevance must be specified to the system (p 8, lines 9 – 14), which is equivalent to the claimed **the statistical information collection step uses content of the one or more retrieved documents** and that **the statistical information collection step considers whether the user-defined predicate has been satisfied by the one or more retrieved documents**, since a determination is made about the ancestors and preprocessed pages are used, which are equivalent to the claimed **one or more retrieved documents**. It should be noted that the *topic* of Chakrabarti et al. is equivalent to the claimed **content** and **predicate**.

26. **Regarding dependent claims 5 and 6**, Chakrabarti et al. teach that we have presented evidence in this section that focused crawling is capable of steadily collecting relevant resources and identifying popular, high-content sites from the crawl, as well as regions of high relevance, to guide itself. It is robust to different starting conditions, and finds good resources that are quite far from its starting point. In comparison, standard crawlers get quickly lost in the noise, even when starting from the same URLs (p 20,

Section 4.8 and p 18, Figure 9), which is equivalent to the claimed **the collected statistical information is used to direct further document retrieval operations toward documents which are similar to the one or more retrieved documents that also satisfy the predicate**, and that **the collected statistical information is used to direct further document retrieval operations toward documents which are more likely to satisfy the predicate than would otherwise occur with respect to document retrieval operations that are not directed using the collected statistical information**, since the focused crawling of Chakrabati et al. utilizes statistical information (p 3) and compares their crawler to other crawlers and outlines the other's shortcomings (Fig 9).

27. **Regarding dependent claim 7**, Chakrabarti et al. teach that multiple citations from a single document are likely to cite semantically related documents as well. This is why the distiller is used to identify pages with large numbers of links to relevant pages (p 8, last paragraph), which is equivalent to the claimed **the collected statistical information is used to direct further document retrieval operations toward documents which are linked to by other documents which also satisfy the predicate**. It should be noted that the semantically related documents of Chakrabarti et al. is equivalent to the claimed **documents which are linked to by other documents which also satisfy the predicate**

28. **Regarding dependent claim 8**, Chakrabarti et al. teach that we describe a Focused Crawler, which seeks, acquires, indexes, and maintains pages on a specific set of topics that represent a relatively narrow segment of the Web. Thus, Web content can be managed by a distributed team of focused crawlers, each specializing in one or a few topics (p 2, fourth paragraph), which is equivalent to the claimed **the information network is the World Wide Web and a document is a web page**.

29. **Regarding claims 10 – 17 and 19 – 26**, the claims incorporate substantially similar subject matter as claims 1 – 8, and are rejected along the same rationale.

***Claim Rejections - 35 USC § 103***

30. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

31. Claims 9, 18 and 27 are rejected under 35 U.S.C. 103(a) as being unpatentable over Chakrabarti et al. as applied to claims 1 – 8, 10 – 17 and 19 – 26 above, and further in view of Chakrabarti et al. (Distributed Hypertext Resource Discovery Through Examples) [as cited by applicant] later referenced as Ch2 et al.

32. **Regarding dependent claim 9**, Chakrabati et al. do not explicitly teach that **the statistical information collection step uses one or more uniform resource locator tokens in the one or more retrieved web pages**.

33. Ch2 et al. teach that other strategies are also known, such as, if the URL is of the form `http://host /path`, then the crawler may truncate components of path and try to fetch these URL's. If links could be traversed backward, e.g. using metadata at the server, the crawler may also fetch pages that point to the page being 'expanded.' (p 382, Column 1, lines 29 – 37), which is equivalent to **the statistical information collection step uses one or more uniform resource locator tokens in the one or more retrieved web pages.**

34. It would have been obvious to one of ordinary skill in the art at the time of the invention to combine the teachings of Chakrabarti et al. with that of Ch2 et al. because such a combination would provide the users of Chakrabarti et al. with teachings of *the architecture of a hypertext resource discovery system using a relational database* (p 375, Column 1, lines 1 & 2).

35. **Regarding claims 10 – 27**, the claims incorporate substantially similar subject matter as claims 1 – 9, and are rejected along the same rationale.

### ***Response to Arguments***

36. Applicant's arguments with respect to claims 1 – 27 have been considered but are moot in view of the new ground(s) of rejection.

### ***Conclusion***

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Nathan Hillery whose telephone number is (571) 272-4091. The examiner can normally be reached on M - F, 10:30 a.m. - 7:00 p.m.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Heather R. Herndon can be reached on (571) 272-4136. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

Doug Hutton  
Primary Examiner  
Art Unit 2176

NH

  
HEATHER R. HERNDON  
SUPERVISORY PATENT EXAMINER  
TECHNOLOGY CENTER 2100